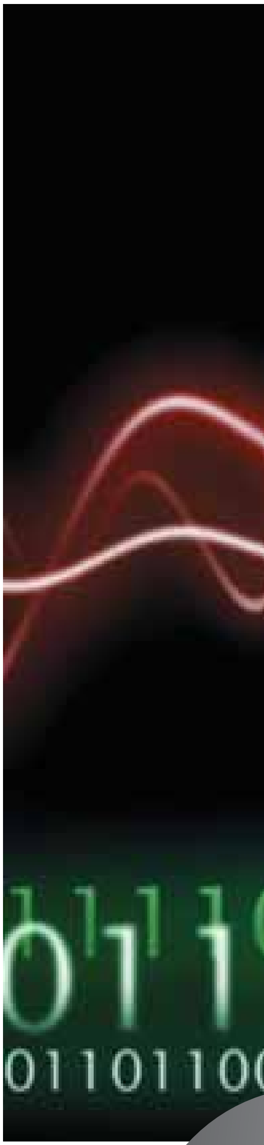
Wireless VoIP

When and Why?

By David Crowe

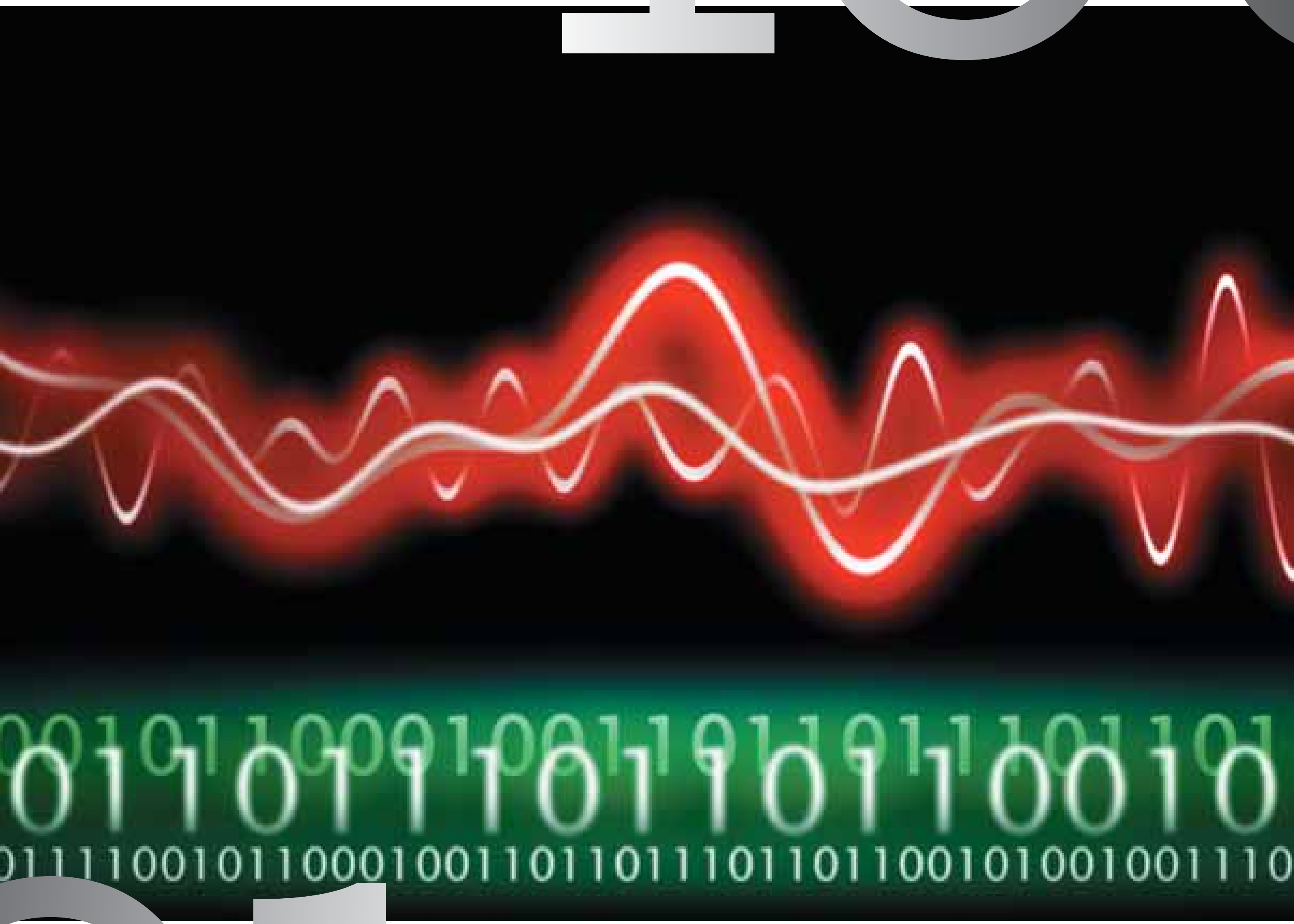


The killer app for public wireless systems might be a surprise to many people. By a country mile, it is still voice – the ability to pick up a phone just about anywhere, anytime and talk to someone else (or sometimes to a machine, but that's another topic). Texting is still growing rapidly, but will probably never overtake voice (except among teenagers). Other data services are growing even faster, but the majority of data users still use voice, and many voice users rarely or never use data beyond texting. In mid-2008, data services in the United States still represented only about 20 percent of total wireless revenues.

But what if voice was data? In one sense, voice already is data. Analog cellular is virtually extinct, so the vast majority of calls are now digitized into a stream of bits inside the phone and are just converted from one digital form to another as they pass between wireless carriers, long-distance carriers and voicemail systems. Only the last mile to wired phone customers is carried as an analog signal.

But in another important sense, voice is not treated like what we normally think of as data – e-mails, Web sites and file downloads. It is carried on separate channels from data when it reaches the wireless carrier, and there is even a separate signaling network using protocols like SS7 and MAP that are unknown in the world of the Internet. Texting is a bit of a hybrid because it also makes some use of the SS7 and MAP signaling networks, but other forms of data, such as e-mail and Web surfing, hop directly onto the Internet from the base station (although they often have dependencies on the signaling associated with the voice network).

To people who see the Internet as the future, maintaining three separate networks – one for voice, one for signaling and one for data – is a crazy idea. Not only crazy, but costly because of the need to manage multiple protocols, and because the equipment for handling voice and signaling systems like SS7 is highly specialized and expensive. IP equipment has lower prices due to economy of scale, and if you look at the cost of pushing a bit through a pipe, this is a dramatic difference.



Voice does not even have high data requirements. Digital voice coders run between 8 and 13 kbps, which is lower than the speed of most dial-up modems. Even on the long-distance backbone, voice calls only occupy 32-64 kbps.

From this knowledge, the concept of wireless Voice-over-Internet-Protocol (VoIP) arose. Why not put everything on the Internet – the voice and the signaling – and mix it in with all other forms of Internet data, e-mail, Web access, streaming video and so on?

This would mean that voice would be transmitted as packets when needed, not as a steady stream. Given that at least half of most conversations is silence (except for people who talk over each other), it seems logical that packetization will increase efficiency, although on the Internet, where bandwidth is cheap, this may not even be a big benefit.

Packetization is one of the fundamental differences between VoIP and first-generation digital telecom. When analog phone systems were converted to digital, the transmission method chosen was Time Division Multiplexing (TDM). Its building block was T1, a single wire (or virtual channel within a larger facility) that is divided into 24 channels (each known as a DS0) which alternate transmitting chunks of 24 voice conversations – each for just under 1/24th of the time. Putting this many calls on a wire that formerly would have carried only one was a huge and obvious advance over analog. Another big advantage was that a digital signal could be regenerated indefinitely without loss of quality and with only negligible delays inserted into the call.

VoIP takes a different approach. Voice is digitized in a similar way, but then the packets are injected into an IP network (not necessarily the public Internet) and mixed with other traffic. The packets may arrive out of order, some may take different routes and some may even be lost. On a lightly loaded network, all the packets will arrive, and on time, but under heavier loads some packets may not arrive or not until it is too late to deliver them.

Here's an impressive word for your next cocktail party – "isochronous." This word, which means "equal time," is a unique and important characteristic of digital voice communications. It means that time slices of voice must be delivered in a regular fashion. Even a few milliseconds of delay ("jitter") can distort the sound enough that it becomes unintelligible.

This characteristic is also required with services like streaming video, where a poor connection will result in the video running jerkily or even pausing while buffering more packets. This can be annoying, but the requirements of a voice conversation are actually more stringent. While video streaming can implement a delay of many seconds at the beginning of playback to fill a large buffer, voice cannot be buffered for more than a few milliseconds because humans react to very quick feedback from the person they are talking to. Anyone who has talked on a satellite phone, where the enormous distances means that the speed of light introduces noticeable delays, will have experienced difficulties which can result in the two people in the conversation to talk over each other.

The isochronous nature of voice means that buffering and retransmissions cannot paper over flaws or overloading in the underlying network.

This is why VoIP and TDM respond so differently to network overload. While an overloaded VoIP system will result in garbled or dropped calls, a TDM network will simply refuse to accept a new call if there are no time slots available at any point in the connection between two parties. This is the fast-busy signal that you may have experienced on Mother's Day or on an overloaded cell site. Once a circuit is established, TDM systems are virtually immune from overload.

VoIP designers have reacted to this in several ways, by simply throwing more bandwidth at the problem, or by introducing Quality of Service (QoS) protocols like RTCP that prioritize voice packets to increase the probability of them arriving on time. It is also necessary to optimize the network design, much like an old analog network, to reduce the number of hops that voice packets take.

Another characteristic of some VoIP systems that appealed to a lot of people was its "end-to-end" nature. Communication of packets can be controlled by the two end points of a conversation, with intermediate points not even necessarily knowing that they are carrying voice packets. Examples of services with this characteristic are Skype and Jeff Pulver's Free World Dial-Up, where two computers or IP phones send packets between them over a generic Internet connection. Other VoIP systems are actually designed to support regular phones, with IP just used as an alternative backbone technology. In other systems, just some of the switching and transmission facilities are IP-based, and much of the network is still circuit-switched.

The end-to-end nature of some VoIP systems led to a significant portion of proponents claiming it should be free, except for the cost of Internet access. Nobody had a right to interfere with the relationship of two consenting phones by inserting arbitrary charges. Sometimes this was a philosophical position, but in other cases it was a product of the dot-com boom where the idea was to build a business first and figure out how to generate revenue later. While the main attraction of VoIP systems is that they are cheaper than traditional systems, the idea that they should be free has gone by the wayside, some during the bust of the dot-com boom, and others because of the recognition that a sophisticated telecom system is a complex business, not just a matter of two devices with an agnostic network in between. Even the free VoIP evangelist Pulver was forced to start charging in September 2008.

Wireless carriers have a wider vision. Not only can they see economies combining their core network into one technology, but it is possible to see the same happening on their air interface. Currently, many systems are based on a voice network to which Internet data channels were an afterthought, but some people believe that future wireless systems will be like Wi-Fi – pure data systems. They will not support voice except as data. This adds another potential efficiency by allowing the same radio interface to be used throughout all spectrum held by the operator, and here VoIP is the only choice. This contrasts with the nightmare of fragmentation that some carriers faced for several years when they had some spectrum devoted to analog, some to one digital technology and some to a more advanced digital technology. A carrier could be out of channels with one technology, with excess capacity in another. Operators that survived this do not want to relive it.

It is extremely important to note that despite all these potential efficiencies, there is no guarantee that VoIP will actually utilize less data per call than current digital cellular with a TDM backbone. This is because the real efficiency benefits of packetization are also combined with a significant increase in the amount of non-payload data. Instead of sending voice packets in highly optimized frames on a traditional cellular interface, with the signaling largely restricted to the internal network, VoIP requires that every voice sample be in an IP packet and that most signaling goes end to end, resulting in larger packets and many additional packets that have to be carried over the air interface to the mobile. Header compression, which removes much of the information that is repeated from packet to packet, is a big help, but it is still questionable whether VoIP will increase voice capacity on wireless systems.

VoIP may also make it more difficult to charge a premium for voice, despite the additional requirements that this special form of data requires (such as prioritizing packets and perhaps putting artificial limits on overall traffic less than the bandwidth that is available). Since carriers might not know when a phone is using VoIP, they could simply charge for voice-grade QoS instead of charging for VoIP, under the assumption that users without will not achieve acceptable call quality, especially at busy times. This might set carriers against third-party VoIP providers who will claim that this is just a ruse to block them from servicing their customers.

Another issue with end-to-end VoIP is security. Because control messages go end to end, any caller can open a signaling connection right through to the person they are calling (who may well not even know them). This is the reason why many corporations limit VoIP to within their PBX, not wanting to risk attacks from malign outside forces via VoIP phones which would break the airlock of their firewalls.

Consumers currently see VoIP mainly as a cheaper alternative to traditional digital telephony. It is interesting to compare the situation with the early days of digital telephony, which was implemented by carriers because of its frequency efficiency, not because of any benefits to the consumer. It was not cheaper and the voice quality was initially no better. My analysis at the time was that it caught on due to an unexpected factor – better battery life. This made digital phones doubly desirable. First, there was no more need to lug an extra battery and charger everywhere. But more importantly, it led to smaller batteries and smaller phones, and soon analog phones started looking like historical artifacts. Manufacturers lost interest in investing in analog, and after a slow start, digital exploded in popularity.

This may happen with VoIP. At present, consumers mainly see cost as its benefit. And if wireless VoIP does not have significantly lower inherent costs to the wireless carrier, wireless VoIP may not even have this consumer benefit. Carriers either have to hope that some big benefit emerges, or they need to encourage the development of innovative applications that are impossible with circuit-switched telephony.

The use of IP to carry some wireless voice traffic is here to stay, but the day when all voice is carried by an end-to-end IP connection is a long way off. There are many security, efficiency and robustness problems to be conquered first. At present, consumers do not see a lot of difference between VoIP and regular telephony, so the impetus will come from cost reductions and efficiencies for the carrier. These are more likely to occur with 4G systems that were designed mainly for data traffic, rather than 3G systems that evolved from voice-only systems. ■

David Crowe is a wireless standards, technology and numbering resource consultant based in Calgary. He can be reached at David.Crowe@cnp-wireless.com.